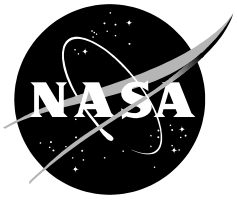
NASA/TM—2010-216395



Measuring and Evaluating Workload: A Primer

Stephen M. Casner, Ph.D
NASA Ames Research Center, Moffett Field, CA

Brian F. Gore, Ph.D.
San Jose State University Research Foundation, San Jose, CA

July 2010

NASA STI Program…in Profile

Since it's founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

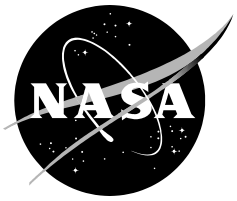
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at http://www.sti.nasa.gov
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA STI Help Desk at (301) 621-0134
- Phone the NASA STI Help Desk at (301) 621-0390
- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

NASA/TM—2010-216395



Measuring and Evaluating Workload: A Primer

Stephen M. Casner, Ph.D.
NASA Ames Research Center, Moffett Field, CA

Brian F. Gore, Ph.D.
San Jose State University Research Foundation, San Jose, CA

National Aeronautics and Space Administration
Ames Research Center
Moffett Field, California 94037

July 2010

Acknowledgements

This research was funded by NASA's Space Human Factors Engineering Project of the Human Research Program (WBS 466199). The authors would like to express sincere appreciation to all reviewers for their input on this document.

Available from:

NASA Center for AeroSpace Information
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

NASA Center for Aerospace Information
7121 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

Table of Contents

Abstract

The workload directed research project surveyed the available literature on: workload measurement techniques; and, the effects of workload on operator performance. Two sets of findings were generated. The first set provided practitioners with a collection of simple-to-use workload measurement techniques along with characterizations of the kinds of tasks each technique has been shown reliably to address. The second set provided practitioners with the guidance needed to design for appropriate kinds and amounts of workload across all tasks for which the operator is responsible. The criterion for inclusion into the EndNote™ reference list database limited the articles to those that were peer reviewed, long standing and accepted in the field, applicable to a relevant range of conditions in a select domain of interest, with parallels being drawn in an attempt to identify analogous "extreme" environments to those in space. Research points towards no one, single approach to determine the suitability of workload in various operational contexts. Different workload evaluations are needed at different phases of the development cycle. The Workload toolbox and the Primer can assist in the selection decision of how and when to use a select set of workload measurement tools.

1. Why Measure and Evaluate Workload?

Human factors researchers have long been interested in the link between workload and human performance. The earliest studies quickly revealed the deleterious effects of workload that is either too high or too low. Humans who are overburdened with work tend to hurry their performance, commit more errors, yield poor accuracy, become frustrated, uncomfortable, and fatigued, and have poor awareness of their surroundings. Interestingly, humans who are underworked can exhibit many of the same symptoms. Low workload has been linked to high error rates, frustration, fatigue, and poor awareness of surroundings as they become bored, as their attention drifts, and as complacency sets in. From all we have learned, it seems that humans do their best when their skills are exercised and their abilities challenged, are neither bored nor overburdened, and when periods of work and rest are equitably mixed together.

These observations about workload and the quality of human performance leave us with two problems.

The first problem is that we need a measuring device for workload. This measuring device would allow us to approach any work situation and acquire a numerical (or at least ordinal) measure of the level of workload being experienced by a human operator. These measures would allow us to determine when person A is working harder than person B or that task A seems to require more work from a human than task B.

The second problem is that we need to define practical and sensible limits for workload. In our discussions of workload, we have tossed around terms such as "overworked" and "underworked." If we are to make scientific judgments about workload levels using numerical measures we collect using our workload measurement device, then we need to more rigorously define these terms.

The ideal solution to both of these problems might be the gauge shown in Figure 1. This gauge displays the level of workload currently being experienced by a human operator against a backdrop of clearly defined limits on acceptable levels of workload. When the human is working "in the green," all is likely well. Excursions into the yellow or red segments of the gauge are likely to trigger some of the ill effects of workload that is too high or too low and result in poor overall performance.

Figure 1. Workload gauge.

It should come as no surprise that human factors researchers have devoted considerable time and effort to developing tools for measuring and evaluating workload that are at least similar in spirit to the gauge in Figure 1. Equipped with tools like these, we could approach any work situation and quickly determine if a combination of human and work environment was operating at peak performance or if adjustments needed to be made in order to raise productivity and performance.

Efforts to develop workload measurement and evaluation tools have ultimately met with limited success and the measurement and evaluation of workload is far from the exact science we would like it to be. There is no workload measurement scale or technique that offers the same reliability as the scales used to measure height, weight, pressure, or other physical quantities. Similarly, our efforts to rigorously quantify the point at which human operators reach their workload "boiling point" fall far short of those that quantify similarly important states of physical matter.

1.1 Challenges to Measuring Workload

In our efforts to measure workload, we will see that the first challenge arises when we attempt to define the very notion of workload. Workload means different things to different people. For example, the word "workload" invites us to think not only about the amount of work that must be performed but also the load or burden that we might experience while performing the work. Some people think of workload as something physical while others believe workload to be more about mental activity or time pressure.

We encounter other challenges when we consider differences in ability and skill among workers. A highly skilled operator might experience a fraction of the workload experienced by another operator

who is performing the same task for the first time. Even comparing two operators at the same experience level, one may have figured out a clever strategy for getting the job done easily while the other operator toils away, doing it the hard way.

Despite the elusive nature of the concept of workload, this Primer will help you in your efforts to measure the compelling concept that we all have: the extent to which we are working hard. Toward this end, we will describe a variety of workload measurement techniques, emphasizing the advantages and disadvantages of each technique for different kinds of tasks and missions.

1.2 Challenges to Evaluating Workload

In our efforts to evaluate workload that has been measured using a workload measurement technique, we face additional challenges. We might adopt a coldly rational approach that seeks to get the highest quantity and quality of work from the human operator in the shortest amount of time. This strategy ignores the health and happiness of the worker who might expire from exhaustion after a few days of record-breaking performance. Similarly, we might focus our measures on the enjoyment or comfort levels experienced by the worker only to find that some workers enjoy not working at all!

We will see how making absolute judgments about levels of workload for individual activities is difficult. An astronaut in the middle of a vigorous physical workout might exhibit signs of high workload only to exhibit signs of low workload while sitting in a meeting later that same day. Looking at the two episodes in isolation, they might seem to exceed the limits on our gauge in Figure 1. Looking at the entire day of activities, we might conclude that the two periods of high and low workload make up a part of a balanced diet of activities. Nevertheless, it is sometimes interesting or necessary to attempt to quantify and judge workload for a single task. In these cases, we will ultimately suggest using several workload measurement techniques at once, looking for collateral agreement between them. Even when we are suspicious about the results provided by a single measurement technique, we are standing on firmer ground when we see that several techniques have produced similar results.

The most practical questions to ask and answer about workload are those that involve comparing the workload experienced by humans when performing several different tasks. In this way, the biases and limitations that apply when measuring workload for one task consistently apply to all tasks that are being compared.

2. Measuring Workload

Four basic approaches to measuring workload have been proposed. We will overview each of these basic approaches, review their advantages and disadvantages, and demonstrate the use of a particular workload measurement technique that uses each approach.

2.1 Performance Measures

Some workload measurement techniques focus on objectively measuring aspects of the operators' performance. Performance measures of workload all have a common characteristic: they consider only the task being performed or the work being produced by the human operator.

2.1.1 Measuring Speed and Accuracy

The simplest performance measurement technique measures the speed and/or accuracy at which an operator is able to perform a task. This approach is founded on the assumption that the operator's performance is likely to degrade as workload increases. Measuring speed and accuracy can be as simple as using a stopwatch to measure the time required to complete a task and noting the operator's success. There are many advantages in using performance to estimate workload. The experimenter needs to do little more than observe the operator performing a task and decide whether or not the operator's level of performance is acceptable. If performance is acceptable, workload is then assumed to be acceptable.

There are also a number of disadvantages of this approach. Measuring only speed and accuracy is rather insensitive to the state or condition of the operator while he or she performs the task. If a human operator feels that he is desperately overworked or underperforming, these observations will not be considered when a performance measure is used. This can also be problematic because when a task is performed for a lengthy period of time the operator becomes fatigued, bored, or falls into any number of unacceptable conditions. For example, a very low workload task might result in excellent performance during the first few minutes of a task but then degrade if the operator becomes bored and his attention begins to drift. Similarly, operators might respond to the challenge of a high-workload task for a short time, turning in a commendable performance, but then begin to fatigue after a time. The following is an example of simple time and accuracy measures.

> *Example:* After the commercial introduction of the typewriter in 1870, researchers began to look at human typing performance (Book, 1908). These studies of typing performance used a simple speed and accuracy measurement technique. Typists were presented with random sentences while the experimenters measured the number of words typed in a given period of time and the number of errors made. It was learned that a typist of average skill could transcribe roughly 60 words per minute using the touch typing technique (using all ten fingers and maintaining eyes focused on the document to be transcribed). A more recent study of computer users showed that today's average computer user is able to transcribe roughly 33 words per minute. "Hunt and peck" typists (who only use two fingers) produce an average of 27 words per minute when copying text (Karat et. al, 1999).

2.1.2 Measuring Activity

Another performance measure that is more sensitive to the state of the operator focuses on measuring the actions that the operator must take in order to complete a task. The basic approach used by this technique is to simply catalog and count the number of steps or actions taken by the operator to complete the task. Large numbers of steps imply high workload, while a task that can be accomplished in only a few steps means low workload. The steps or actions that are counted using this technique might include control inputs, verbal responses, mental calculations, decisions, and gazes or visual searches required to complete a task.

> *Example:* When the first GPS receivers became commercially available, researchers in England used an eye-tracking device to record the glances made by drivers who used two types of navigation information: a paper map and an LCD display. The eye-tracking device captured both the number and duration of glances made by drivers who used each navigation method. The researchers were interested not only in which navigation method drew more glances from the drivers but also in the effect on how many glances were made to each of the vehicle's mirrors and out of the front window (Fairclough, Ashby, & Parkes, 1993). The researchers found that as

the number of glances that were directed toward the LCD screen increased, the number of glances to the rear view and side mirrors decreased.

One advantage of the technique of measuring operator actions is its simplicity: one only need to observe the user's actions, record them as they occur, and tally them up once the task is complete.

A basic disadvantage of this technique is that it does not directly address the notion of workload as many understand it. The finding that a task requires a small or large number of steps to complete does not necessarily mean that the operator will experience a feeling of being underworked or overworked. This approach to measuring workload also has the disadvantage of ignoring skill differences between operators. One operator may effortlessly perform a task that requires great effort from another operator.

2.1.3 Task Analysis

A variation on the technique of tabulating the operators' actions is to enumerate the procedural steps that are required to complete a task—without actually observing operators while they perform the task.

> *Example:* A group of researchers analyzed the steps required to complete a customer assistance task performed by telephone operators at a major telecommunications company. The researchers enumerated the visual searches, button presses, speech utterances, and mental calculations that were required to perform the customer assistance task using the company's existing workstations and then repeated the analysis for operators who used a new workstation that the company was considering. The company was considering the new workstations in hopes that they would help shorten task times and save the company money. Contrary to expectation, the task analyses predicted that the new workstations would *increase* the time required to complete the task. Subsequent empirical testing revealed that the predictions made by the task analysis were indeed correct (Gray, John, & Atwood, 2002).

The main advantage of the task analysis technique is that it requires no testing with human operators. Some disadvantages of the technique include: cataloging the procedural steps can be time-consuming; the technique ignores differences in skill between operators, although some variability can be reduced if the analysts work in the context of a common model or architecture (e.g. GOMS; Gray, John, & Atwood, 2002); and it assumes that all operators will follow the same procedure to complete the task.

2.2 Indirect Measures

An indirect way of measuring workload is to estimate the level of workload imposed by a task by measuring how well operators are able to perform a second task at the same time they are performing the primary task. In this way, workload is estimated by measuring how much "spare capacity" the operator has. If the operator is able to handily perform a second task at the same time as the primary task, then we can conclude that the primary task burdens the operator with only a low or moderate amount of workload. On the other hand, if performing a secondary task leads to a breakdown in the operator's performance of the primary task, we can conclude that the primary task absorbs most of the operator's resources and that the operator is nearing the peak of his capacity to do work.

Example: A recent example of studies that look at the effects of secondary tasks on primary tasks are those that examine the effects of mobile phone conversations on driving performance (Strayer, Drews, & Crouch, 2006). The researchers found that the tasks of producing and comprehending language draw considerably away from the faculties used during the driving task and significantly degrade driving performance.

A number of secondary tasks have been proposed over the years and there is general agreement that there is no single best secondary task (Ogdon, Levine, & Eisner, 1979). However, researchers have provided good advice about selecting a secondary task. First, a secondary task should use the same resources as the primary task. For example, a primary task that requires the operator to monitor events out a window or on a computer display, an ideal secondary task might require the operator to monitor another display for occasional alerts or messages. If the operator consistently misses alerts on the secondary display, it is likely that the primary monitoring task is usurping much of their attentional capacity. A poor choice of a secondary task for this primary task would be to ask the operator to perform mental arithmetic since these mental calculations might be performed simultaneously with the primary monitoring task. Second, a secondary task should require substantial effort to complete. A secondary task that is too easy might fail to interfere with the primary task at all and thus fail to reveal how much work is being performed (Fisk, Derrick, & Schneider, 1983). Gawron (2008) describes 29 secondary tasks for consideration, offering advantages and disadvantages of each task.

For primary tasks requiring visual attention, good choices of secondary tasks are:

- *Sort playing cards:* give operators a deck of shuffled playing cards and ask them to sort the cards by suit and rank (Lysaght et. al, 1989)
- *Detection:* ask operators to monitor a display for occasional alerts

For primary tasks that require mental processing, good choices of secondary tasks might be:

- *Mental math:* ask operators to perform simple mental arithmetic

For primary tasks that require psychomotor skills, this secondary task might be appropriate:

- *Tapping:* ask operators to tap out a number or pattern of sounds with their finger
- *Classification:* ask operators to place a word, number, or object into a category or class (e.g., red is a color, 7 is a number)

Secondary tasks can be combined with primary tasks that require operators to use more than one of their faculties.

Example: Green and Flux (1977) asked pilots who flew a flight simulator to add 3 to a number that was given verbally. The time to respond to the mental addition task was measured and recorded. The results showed that the time to respond well correlated with rises and falls in workload introduced in the flying task. Huddleston and Wilson (1971) obtained a similar result when they asked pilots to determine if a number, or the sum of two numbers, was odd or even. These studies suggest the usefulness of simple mental mathematics exercises as secondary tasks for measuring workload.

Indirect workload measures have the advantage that they offer more clues about the condition of the operator during the performance of a task.

There are a number of disadvantages of indirect measures of workload. The first disadvantage is that they rely on assumptions about which kinds of secondary tasks compete with the performance of the primary task. For example, a secondary task that requires the operator to visually scan a separate display might be an excellent measure of spare attentional capacity for an operator who must monitor a scene through a front window. Diminished performance on the secondary task might allow us to correctly conclude that the primary monitoring task requires much attention on the part of the operator. Unfortunately, this does not allow us to conclude that the primary task is causing the operator to experience extreme levels of workload. It may be the case that the operator could perform other concurrent tasks that do not require frequent and sustained visual attention.

A second disadvantage is that the operator may perform well on the secondary task only to find that their performance of the primary task has been compromised. This can happen when operators turn their attention more to the performance of the secondary task. Note that it is not always apparent to the experimenter when operators decide to neglect the primary task in favor of the secondary task. Another disadvantage lies in using the same secondary task to compare two different primary tasks. It can never be known with certainty how a secondary task overlaps with any given primary task. A fourth disadvantage is that different operators can have different skill levels, or use different strategies to perform either the primary task, secondary task, or the combination of the two tasks.

2.3 Subjective Measures

Subjective workload measures ask the human operator to describe the workload they experience when performing a task. Subjective workload measures do not attempt to measure anything about the task that the user is performing or the user's performance at any task. Subjective workload measures focus entirely on the human operator's feelings about their workload.

Two basic varieties of subjective workload measurement techniques have been developed:

- *Subjective numerical measurement techniques* ask the human operator to assign a numerical or ordinal value to the workload that they are currently experiencing while working in a particular task situation.
- *Subjective comparative measurement techniques* ask the human operator to make comparisons between two or more tasks situations and say which situation results in the higher (or lower) workload.

We will give examples of both types of techniques and explain how each offers its own unique combination of strengths and weaknesses.

2.3.1 Subjective Numerical Measurement Techniques

Instantaneous Self-Assessment

The simplest and least intrusive subjective numerical workload measurement technique is one in which you ask your subjects to rate their overall workload, at periodic intervals, on a scale from 0 to 100. The main advantage of the Instantaneous Self-Assessment (ISA) technique is that it is among the simplest measures to collect. The experimenter need only be equipped with paper and pencil while they observe and query the operator as they perform the task.

A principle disadvantage of the instantaneous self-assessment technique arises from differences in the way people think about workload. For example, some people may regard a task such as carrying boxes up five flights of stairs as a high workload task. For these people, physical labor epitomizes the idea of workload. For others accustomed to a stressful job such as trading on the floor of the New York Stock Exchange or after months of planning a wedding, a day spent carry boxes might be perceived as a welcome relief, perhaps even a form of relaxation. To others accustomed to working under tight deadlines, these types of mental or physical work may not seem like high workload at all. For them, having to complete a task quickly means high workload, regardless of the nature of the task. These differences in the way people perceive workload can result in drastically different subjective workload measures across operators who perform the same task.

Another principle disadvantage of the instantaneous self-assessment technique is that different operators tend to use different portions and different amounts of the 0 to 100 scale. For example, some operators naturally associate 50 with a situation in which they are neither overworked nor underworked and drift linearly toward the ends of the scale as their perceived workload rises and falls. Other operators tend to crowd themselves into one segment of the scale or another, making disproportionate movements to either side. When crowded to one side of the scale or the other, operators sometimes inflict ceiling or floor effects on themselves when perceived workload changes considerably but when they have essentially run out of room at one end of the scale. Attempts can be made to calibrate operators and prompt them to think of a value of 50 as being the middle, and making proportionate movements to both sides, but these attempts have typically met with limited success. For a number of reasons, operators think of workload and their experience of workload in unique and personal ways. Personality factors have been associated with atypical usage of the 0 to 100 scale (Hart, personal communication). Operators who feel that they are highly skilled can sometimes cling to the lower end of the scale, indicating that their formidable skills are only partly tapped by any task.

Researchers have also questioned to what extent operators' impressions of workload are colored by their perceptions of the quality of their own performance. That is, when operators feel they are performing at a substandard level, they may also feel their workload is high.

NASA Task Load Index

The NASA Task Load Index (TLX) measurement technique was developed to help mitigate a number of problems that arise from differences in the way people think about workload. The NASA TLX technique is similar to the instantaneous self-assessment technique in that the experimenter must periodically ask the human operator for subjective estimations of his/her workload. The key difference about the TLX technique is that, rather than asking participants to subjectively rate their workload using a single scale, participants must subjectively rate their workload along six different workload sub-scales. Each of the six workload sub-scales was designed to characterize workload in a different way. The six workload sub-scales are as follows:

1. Mental Demand
2. Physical Demand
3. Temporal Demand
4. Performance
5. Frustration
6. Effort

The idea behind using these six different workload sub-scales is to obtain at least one measure of workload from each participant that captures the essence of workload—the way they conceptualize it in their mind. For example, if a participant feels that working under time pressure epitomizes workload, then the estimate they provide for the Temporal Demand might best capture his/her level of workload—as he/she sees it. In this way, the six workload sub-scales work together to accommodate six different ways of thinking about workload.

Using TLX, operators are asked to provide ratings along each of the six workload sub-scales. These ratings can be provided verbally to an experimenter or by using any other data collection device including paper and pencil or a computer.

Collecting the six measures from each participant leaves the problem of combining the measures to arrive at some overall measure of workload. The NASA TLX technique requires participants to rank-order the six workload sub-scales in terms of which sub-scale, in their minds, better characterizes workload.

A seventh measure of workload (overall workload) is then calculated by multiplying each workload rating by its numerical ranking (1 though 6), adding these weighted rankings together (1+2+3+4+5+6), and then dividing by 21 to arrive at the overall workload measure. When all is finished, the technique yields seven measures of workload: the six individual measures plus the overall workload measure. It is usually interesting to look at, evaluate, and compare all seven workload measures.

> *Example:* Casner (2009) used NASA TLX to examine differences in pilot workload in two types of airplanes: one outfitted with conventional navigation equipment and instruments, the other outfitted with advanced navigation and control equipment. Pilots flew from airport to airport, passing through four distinct phases of flight: Setup, En Route, Approach, and Missed Approach. The experimenter used the last 30 seconds of each flight phase to verbally collect the pilots' workload ratings along each of the six TLX workload sub-scales. Ratings were recorded using the score sheet shown in Figure 2 as the pilots verbally provided the ratings.
>
> At the end of the flight, pilots were asked to rank the six TLX sub-scales in terms of which best characterized the notion of workload in their minds. Back in the lab, the workload ratings and the rankings were used to calculate the overall workload measure and to determine if any differences between the two airplanes existed. The results showed that the advanced navigation and control equipment helped lower workload during some phases of flight but raised workload in other phases. Overall, there was no difference between the two airplanes.

Among the advantages of the TLX technique is that TLX is more accommodative of different ways of conceptualizing the notion of workload. TLX offers the flexibility of collecting workload measures while participants perform the task or after completion of a task while the operator's memory of the task experience is still fresh. Workload ratings can be collected from participants verbally, using a pen and paper, or by computer interface. This flexibility allows NASA TLX to be used for tasks in which the participant's eyes are free or for tasks in which the participant must remain "heads up." The TLX technique also attempts to accommodate any biases about workload that might stem from operators' perceptions of the quality of their own performance.

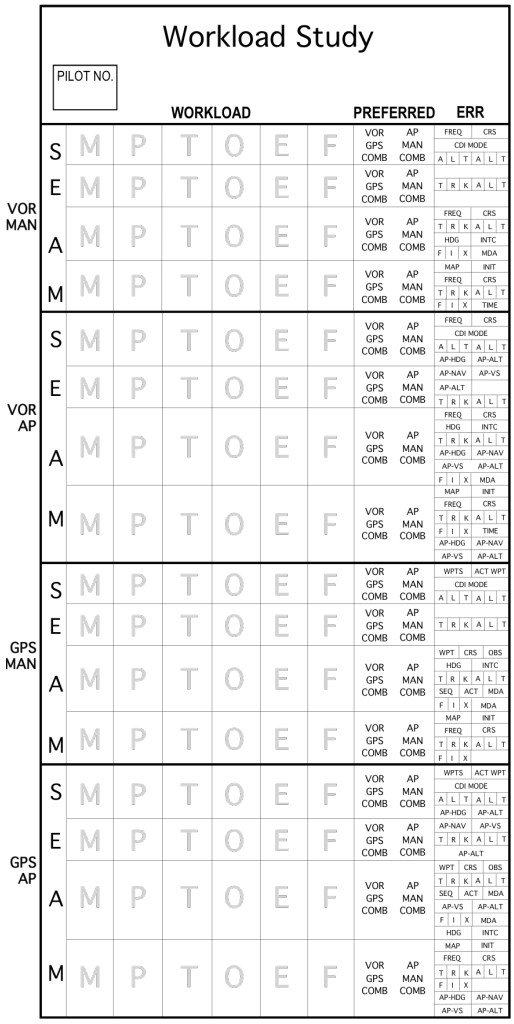


Figure 2. NASA TLX score sheet.

Among the disadvantages of the TLX method is that TLX is more time-consuming than other techniques since measures along six different sub-scales are required. The TLX technique suffers from the same "scale loading" problems that the ISA technique does: operators do not always think of a value of 50 as the middle and move linearly toward the two ends of the scale as perceived workload rises and falls.

Bedford

The Bedford Workload Scale also collects subjective ratings of workload from participants. The Bedford technique presents the operator with a 10-element scale and offers some of the simplicity of the Instantaneous Self-Assessment technique. In an attempt to circumvent the scale-loading problems associated with the ISA and TLX techniques, the Bedford scale attaches elaborate verbal descriptions to each of the 10 values along the scale, as shown in Figure 3. To simplify the process of choosing one of the ten workload ratings, the Bedford scale juxtaposes a hierarchical decision tree onto the ten scale ratings. Operators must navigate through the hierarchy and narrow down their choices of workload ratings to two or three choices and then select a single rating based on the descriptions attached to the ratings.

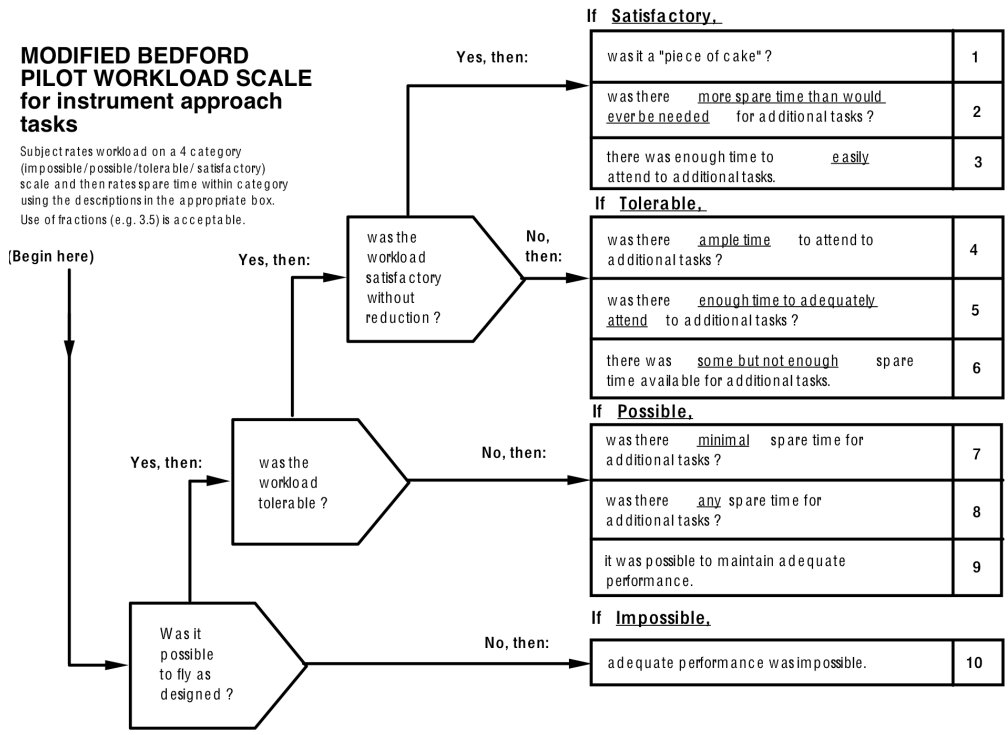


Figure 3. The Modified Bedford Pilot Workload Scale (for instrument approach tasks)[1].

An important advantage of the Bedford technique is that it associates descriptions with each of the values along the 1 to 10 scale. Another advantage is that the descriptions themselves represent interpretations of the ratings offered by operators. That is, if an operator offers a rating of 7 on the Bedford scale, the text description that is associated with that rating provides its own interpretation of the rating.

One disadvantage of the Bedford technique is that it often can only be used after each participant has completed a task or when the participant's eyes are free to focus on a paper or display that shows the Bedford scale. Another limitation of the Bedford technique is that the ten choices on the scale do not comprise an interval scale. That is, although numbers are assigned to the ten choices, the distance between the ten choices cannot be construed as equal (i.e., a rating of 6 does not represent twice the workload of a rating of 3). Another limitation of the Bedford scale is that as operators become proficient with the scale, they report they no longer use the hierarchical choices and proceed directly to the ten ratings. Lastly, and probably most important, the Bedford scale asks subjects to make judgments about the notion of "spare capacity." Similar to the ambiguities introduced by presenting the work "workload" to subjects, the phrase "space capacity" can be interpreted as the situation in which the operator has additional time, additional mental capacity, a free hand, etc. An operator that thinks of spare capacity in one of these ways might give vastly different ratings than an operator who thinks of spare capacity differently.

[1] Roscoe, 1984

Advantages and Disadvantages of Subjective Numerical Workload Measurement Techniques

Although subjective workload measures have been criticized as being less "scientific" than other types of workload measures, they have a compelling and almost irrefutable face validity. As one researcher points out, if an operator complains about being overworked or underworked, this is likely to be the case regardless of what any other types of measures may show (Moray et. al, 1979). Regardless of their apparent validity, subjective workload measures introduce their own unique risks when they are used to make absolute judgments about operator workload.

Since different operators tend to load onto numerical scales differently and think about the very notion of workload differently, subjective workload measurement techniques tend to yield consistently moderate average workload ratings together with large variability. For example, in a recent study of experienced pilots using a GPS navigation computer and an autopilot in an advanced cockpit airplane, eight pilots provided the subjective workload ratings for the Mental and Temporal Demand sub-scales shown in Figure 4.

All eight pilots performed the same task using the same equipment under the same conditions. Considered together, the mean workload ratings work out to a comforting value that suggests that operators are neither underworked nor overworked. However, looking at the individual scores, these conclusions become suspect. The individual measures suggest that Pilot 1 is grossly underworked while Pilots 2 and 6 approach what might be considered an upper limit on workload. When a confidence interval is calculated using the individual scores, we can say that there is a 95% chance that the true average workload falls between a value of 26 and 56. All in all, the data might allow us to conclude that these pilots were unlikely to be overworked but not to say with any degree of confidence whether workloads were acceptable or too low. Looking at the two measures provided by each pilot, there is some consistency. The consistency in the measures along the two sub-scales might suggest that the two pilots think differently about the 0 to 100 scale but also introduces the possibility that they experienced two drastically different amounts of workload.

Pilot	Mental Demand	Temporal Demand
1	5	5
2	65	70
3	45	50
4	30	65
5	30	35
6	75	60
7	40	50
8	40	40
Average	**41.3**	**46.9**

Figure 4. Example of NASA TLX Workload scores for eight pilots.

Another disadvantage of subjective workload techniques is that they do not always measure workload as the operator performs a task, but rather sometimes after the task has been

completed. This has prompted a number of claims that subjective measurement techniques are a better test of the operator's memory than they are of the workload they had experienced (O'Donnell & Eggemeier, 1986).

2.3.2 Subjective Comparative Measurement Techniques

Subjective Workload Dominance Technique

The subjective workload dominance (SWORD) technique begins with the assumption that a more reliable evaluation of the workload experienced by human operators in various task situations can be achieved when the task situations are compared with one another rather than evaluated in absolute terms. For example, if we are interested in knowing how tasks A, B, and C compare with respect to workload, the SWORD technique asks operators to make comparative judgments between each combination of the three tasks (A vs. B; B vs. C; and A vs. C). Comparisons are solicited from operators using a comparison sheet such as the one shown in Figure 5.

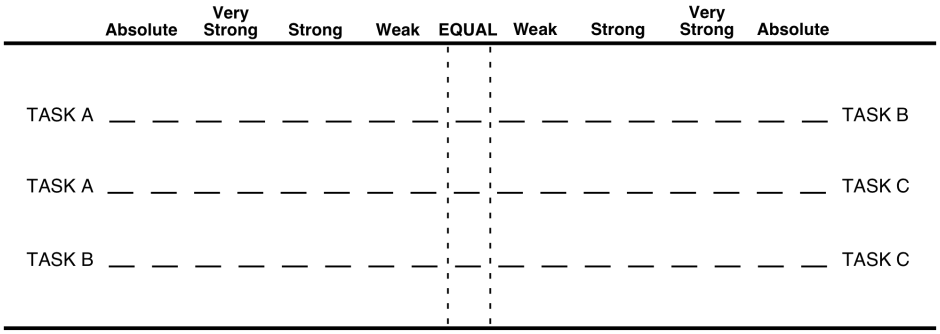


Figure 5. SWORD response sheet.

Operators make relative comparisons using a 17-element scale. A mark on the middle element indicates that the subjective workload experienced while performing the two tasks is roughly equal. Marks made to the right or to the left of the middle element indicate that workload was greater for one task or the other to increasing degrees.

The developers of the SWORD technique recognized that the relative comparisons made by operators may not always be consistent (e.g., operators might not say that workload for task A was four times greater than task C, even though they stated that A was twice as great as task B, and B was twice as great as task C). For this reason, the comparisons made by operators are subjected to a statistical analysis that evaluates the degree of consistency among the rankings and assigns a reliability estimate of the entire workload evaluation based on the degree of consistency. A number of statistical techniques have been proposed and used for determining the reliability of the estimates provided by operators (Budescu, Zwick, & Rapoport, 1986; Crawford & Williams, 1980).

Advantages and Disadvantages of Subjective Comparative Workload Measurement Techniques

A principle advantage of subjective comparative techniques is that they do not require operators to assign numerical or ordinal rankings to the workload they experience. This eliminates the problems associated with scale usage and the problems associated with interpreting key words and phrases that might be used to describe numerical levels of workload.

A main disadvantage of comparative techniques is that they offer few means of discovering when any given task situation results in too much workload. However, we will defer the discussion of workload limits to a later section.

2.4 Physiological Measures

Physiological measures of workload attempt to associate physiological changes with levels of workload. For a time, researchers hoped that physiological measures could be found that represented a truly objective workload measurement—one that did not rely on assumptions about how people perceive workload or on subjective ratings provided by human operators. For this hope to come true, a physiological measure would have to be found that represents a "tell tale" sign of the experience of workload. While many physiological measures have been investigated, no one measure has proven to definitively capture our notion of workload.

2.4.1 Heart Rate

Perhaps the simplest and most time-honored physiological measure of workload is heart rate. Heart rate can be measured using a simple heart rate monitor such as those used during exercise. These devices sample and record heart rate roughly once per second and store the data in files that can easily be viewed and analyzed on any personal computer. Numerous researchers have studied the extent to which heart rate correlates to human task performance and to other measures of workload and the results have been mixed. Heart rate seems to be a good correlate of physical activity and a modest correlate of mental activity. Roscoe (1992) reviews a number of studies in which heart rate was used to measure workload among pilots. The earliest of these studies dates back to 1917.

2.4.2 Heart Rate Variability

A slightly more sophisticated measure of workload is heart rate variability. Heart rate variability is the differences in the time intervals between heart beats, irrespective of the number of beats per second. Measuring heart rate variability requires more sophisticated equipment. A number of researchers have had some success in relating heart rate variability with other measures of mental workload (Vicente, Thornton, & Moray, 1982; Mulder, 1980; Metalis, 1991).

2.4.3 Evoked Potentials

More sophisticated physiological recording techniques measure changes in electrical potential in responses to visual and auditory events or capture images of the brain while operators perform tasks. These highly specialized (and costly) measurement techniques are beyond the scope of this Primer.

2.4.4 Advantages and Disadvantages of Physiological Measures

A primary advantage of physiological workload measures is their unobtrusiveness. Unlike workload measurement techniques that require operators to perform secondary tasks or to provide verbal estimates of their own workload, physiological measures tacitly measure biological changes in the operator as they work.

A primary disadvantage of physiological workload measures is that there is little theory behind them. Although we have observed and associated changes in the cardiac, respiratory, and central nervous systems while operators work and can make sensible hypotheses about why these changes might occur, there is no clear-cut mechanism by which the same physiological changes should occur in every operator as they perform work. As is the case with the other types of

workload measurement techniques, researchers have had mixed success in relating physiological measures to workload.

3. Evaluating Workload

Researchers and developers use workload measurement techniques because they have questions they need to answer. Before selecting a particular workload measurement technique and getting started, it is important to establish the questions you wish to answer. Two questions about workload seem to naturally arise in the minds of practitioners who become interested in measuring workload.

3.1 Is Workload Too High or Too Low?

A natural question to ask in any task or mission situation is: "Is the operator's workload too high or too low?" Since most workload measurement techniques yield some sort of quantitative measure (subjective comparative techniques such as SWORD are an exception), this question leaves us with the problem of interpreting these measures in absolute terms. For example, if a workload measurement technique yields a workload measure of 70, we must decide whether 70 is acceptable, too high, or too low. Making this judgment introduces the need to define acceptable ranges of operator workload. Taking this idea further, it would seem useful to be able to define absolute upper and lower limits on workload: the equivalent of redlines at the extremes of our workload gauge shown in Figure 1. Reflecting on the limitations of each type of workload measure technique discussed above, we encounter difficulties when trying to interpret workload measures literally or to make judgments about workload measures in absolute terms. Although some researchers have proposed redline limits for particular workload measures (Rueb, Vidulich, & Hassoun, 1994; Reid & Colle, 1988) for these reasons, efforts to define redline values for workload have met with limited success.

Direct workload measurement techniques that measure speed and accuracy are problematic because they fail to identify high workload as the reason for poor performance. It is a questionable assumption that high workload is likely to result in poor performance (Yeh & Wickens, 1988; Vidulich & Wickens, 1986). Direct workload measurement techniques that measure or model operator activity also fail to consider the condition of the operator while they perform the activities and to consider whether or not the operator is underworked or overworked.

Indirect workload techniques can be unreliable when the relationship between primary and secondary tasks is unclear. Variations in subjective workload ratings provided by different operators tend to produce moderate overall workload ratings when they are aggregated across individuals.

Subjective workload measurement techniques introduce the problem of operators using different portions and ranges of the scale. Even when normalization techniques are used to attempt to place all operators' ratings on a common scale, we are still left with the problem of some operators not revealing states of being overworked.

The idea of defining a redline limit for workload raises another question: what is the meaning of redline? Rueb, Vidulich, & Hassoun (1994) drew an analogy between a workload redline and the redline shown on the tachometer of an automobile. The tachometer redline indicates a value at which sustained performance will likely result in harmful effects, not a value at which the engine will instantaneously fail. This idea poses the challenge of defining a time limit to accompany the

workload limit: how long can human operators spend at or above the redline? Although the negative effects of prolonged periods of high or low workload are well documented, there is good evidence to support the idea that periods of hard work, followed by periods of rest, are beneficial. Regular practice in using skills associated with emergency procedures or any fast-paced or demanding work situations can only help improve operator preparedness. One way to circumvent this problem is to introduce a notion of *workload dose* that is similar to the notion of noise dose used for noise exposure. Rather than placing all emphasis on peak levels of workload, a workload dose approach would consider episodes of workload individually, noting both the level and duration of the workload. These individual episodes can be summed to arrive at an overall workload dose.

Operator workload is often an item of interest in complex situations in which humans perform many different tasks as part of an ongoing mission. For example, we might be potentially interested in how workload changes throughout an entire day of work and as operators perform dozens of individual tasks. This presents the question of on which tasks should workload be measured.

One solution is to measure workload throughout the entire mission. Most of the workload measurement techniques described above can be used to collect periodic measurements throughout the course of a mission. The resulting data can be used to determine how workload rises and falls over the entire work period.

An alternative approach to answering questions about absolute workload is to use several different workload measurement techniques concurrently (Hendy, Hamilton, & Landry, 1993). If several techniques yield results that are in agreement with one another, then there is collateral evidence that the findings are representative of operator workload. Regardless, practitioners are cautioned against using any numerical workload measure as an absolute measure of workload. While researchers who developed workload measurement techniques may have originally had the goal of creating a measuring device for workload that had the same interval properties as scales used to measure physical properties, workload scales have not yet achieved such a status.

Another solution is to measure workload for tasks that are identified as being particularly important. There are a number of criteria for choosing which tasks are most important for workload measurement:

- tasks for which the possibility of error is less acceptable
- tasks for which no oversight or redundancy is available
- tasks for which workload is suspected or reported to be high

3.2 How Does Workload Compare between Several Tasks or Designs?

Because of the limitations inherent in all workload measurement techniques, a more practical goal in assessing workload might be to compare the workload measures obtained for one task or design with those obtained for other tasks and designs. The reason for this is simple: even if operators think differently about workload, have different skill levels, use different strategies, load onto scales differently, or experience different physiological responses, it is likely that they will carry these same differences from task to task and design to design. For example, if an operator consistently provides an average workload rating of 60 for Task A and an average workload rating of 40 for Task B, then we are on more solid ground in concluding that operator is experiencing higher workload on Task A. There is a long history of using a single workload measurement technique for making these kinds of workload assessments and design decisions based on them.

3.3 Undertaking a Workload Measurement and Evaluation Effort

We have described the tools needed to conduct an analysis of workload for most situations in which humans work alone, with others, or together with technology to accomplish an aim. We now turn our attention to the steps needed to undertake such a study: from the formulation of the questions that one wishes to answer,to the drawing of conclusions. We will not attempt to review the entire process of conducting a scientific experiment here but rather review those parts of the process that are specific to measuring and evaluating workload.

The use of the workload measurement and evaluation techniques we have described generally requires five steps.

3.3.1 Formulate Questions

The first step is to formulate the questions that you wish to answer using workload analysis. While everyone may have different questions to answer, common questions about workload might include:

- Does workload for one task differ from workload for another task?
- What level of workload do operators experience for a task?
- Does workload vary between operators?
- Does workload vary through different phases of a task?
- What is the highest, lowest, or average workload experienced by operators for a task?

3.3.2 Choose Workload Measurement Technique(s)

Based on the kinds of questions you wish to answer and the details of your task environment, your next step is to choose a workload measurement technique. The table in Figure 6 summarizes the characteristics of each workload measurement technique described above. The table makes explicit how workload measurement techniques are different in the amount of preparatory work they require; how much time they require from the operator; and whether or not they require visual, auditory attention, or manual intervention from the operator. The table in Figure 6 also summarizes how the techniques vary in the type of measurement they provide: absolute measures; absolute measures that are anchored to a particular scale; or relative measures.

Technique	Prep	Time	Eyes	Ears	Hands	Absolute	Relative	Anchored
Performance								
Speed and Accuracy						x		
Activity						x		
Task Analysis	x							
Indirect								
Secondary Tasks		x	x	x	x	x		
Subjective								
Numerical								
ISA				x		x		
TLX		x		x		x		
Bedford			x	x		x		x
Comparative								
SWORD			x		x		x	
Physiological								
Heart Rate						x		
HRV						x		
Evoked Potentials						x		

Figure 6. Characteristics of workload measurement techniques considered.

3.3.3 Prepare Materials for Data Collection

The next step in preparing to use a workload measurement technique is to create a means of measuring and recording the workload data. If workload measures are to be made through observation by an experimenter, a simple log sheet can often be made that simplifies the recording of measures. For example, Figure 7 shows a simple score sheet that allows an experimenter to record performance times for tasks and record the number of instances of errors of three particular types for a simple tracking task.

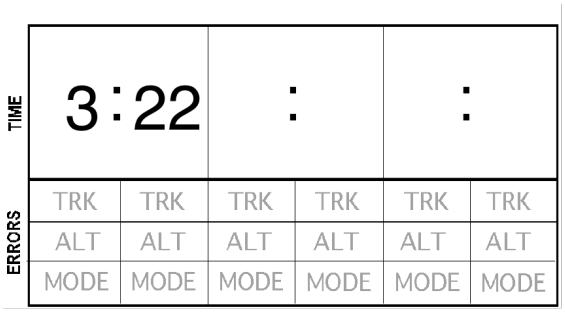


Figure 7. Simple log for operator performance time and error on a tracking task.

In the case that a subjective workload measurement technique is used, score sheets can be created that allow an experimenter to record ratings that are given verbally by operators. A simple score sheet that allows an experimenter to collect instantaneous self-assessment, TLX, and Bedford measures is shown in Figure 8.

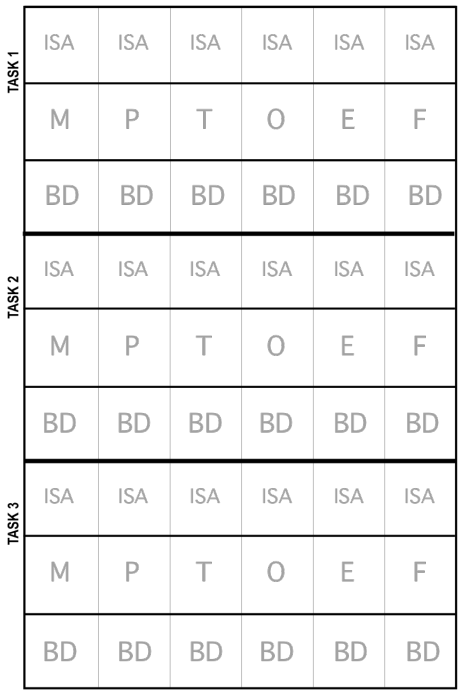


Figure 8. ISA, TLX, and Bedford rating sheets used by experimenters.

Note that in order to use the Bedford technique, operators must also be provided with the Bedford scale, shown in Figure 3, as a reference.

Subjective workload rating techniques are often used in situations in which it is best to have the operator directly record their own workload measures. In these cases, some type of computer apparatus is required. In the case of some techniques such as TLX, there are accompanying software packages or Internet sites that allow operators to enter workload ratings directly after they have completed a task. Other experimenters have created their own custom software and hardware tools that allow operators to directly enter workload ratings while they perform a task. These tools might present the operator with a message on a computer screen indicating that a workload rating is now requested. The operator would then respond to this request by pressing a numerical key on the computer keyboard.

3.3.4 Choose and Instruct Participants and Gather Data

After a workload measurement technique has been chosen and the needed materials have been prepared, individuals must be chosen to participate in the workload measurement experiment. After participants are chosen, they must be instructed on the use of each workload measurement technique that requires action or input from the participant. The ISA, TLX, and Bedford techniques all require subjective workload ratings from participants and therefore require that participants be trained in the use of each technique. Participants should be given the basic instructions that are provided in the seminal publication that describes each technique. All of these instructions provide participants with a basic familiarization of the concept of workload, outline what responses (ratings) will be required from participants, and often include some basic description of the workload scales to be used. For example, techniques that use a numerical 1 to 100 scale sometimes instruct participants to think of 50 as the middle of the scale (the situation in which the participant feels neither underworked nor overworked).

Of particular importance is providing participants with an opportunity to practice giving subjective workload ratings. A number of authors have noted how participants' subjective workload ratings can vary as they acquire more practice with using the scale.

Of equal importance is providing the experimenter(s) with a chance to practice whichever data collection steps they must perform as the experiment progresses. Even if these steps only require the experimenter to note important events and record them, or verbally acquire subjective workload ratings from operators and record them, it is important that the experimenter reach a point of proficiency in performing these steps.

3.3.5 Analyze Data and Draw Conclusions

Once the data have been collected, they must be analyzed and used to answer the questions that were posed.

3.4 An Example: Using a Flight Management Computer

We will illustrate the process of conducting an analysis of workload using a hypothetical example. For this example, we will first establish the question or questions we wish to answer about workload. We will then choose one or more workload measurement techniques based on the questions we wish

to answer and the details of the tasks we wish to study. Then we will walk through the process of using the workload measurement technique(s) to gather data with test subjects. Finally, we will analyze the data we gathered to help answer our questions.

3.4.1 Task to be Studied

Flight management computers that incorporate GPS technology and the ability to pre-program flight routes are now common in aircraft of all types. The flight management computer shown in Figure 9 is common in general aviation aircraft.

The flight management computer allows the pilot to pre-program a flight route and then receive guidance along that route during flight. As long as no changes are required during the flight, following the route amounts to little more than proceeding from one waypoint in the route to the next. Figure 9 shows a simple route that has been preprogrammed. The waypoints that comprise the programmed route are shown as a list of waypoints at the bottom right of the display.

Figure 9. Flight management computer used in general aviation.

It is sometimes the case that air traffic control will ask pilots to make changes to the programmed route during flight. This situation is often cited by pilots as being one that raises workload to undesirable levels. In order to change the flight route, pilots must divert their attention from the out-the-window scene to the flight management computer. Since a series of button presses and menu selections are required to effect the changes in the route, a period of "head-down" time is required. When these changes are requested by air traffic control during the approach phase of flight (nearing the destination airport), workload is often cited to become high since pilots must complete several other tasks during this busy phase of flight (radio communications, briefing approach procedures, etc.).

3.4.2 Questions to be Answered

In this example, we will walk through the process of doing a simple workload analysis of the route-changing task using the flight management computer shown in Figure 1 during the approach phase of flight. In our example, we are mainly concerned with discovering if workload becomes what might be regarded as too high.

3.4.3 Choosing Workload Measurement Techniques

In this example, we are interested in making an absolute judgment about workload levels. Considering the limitations of all workload measurement techniques for drawing these sorts of conclusions, we decide to use several measurement techniques and look for collateral evidence that workload might in fact be too high. Since performing the task in an error-free fashion is of the utmost importance, we decide to include a direct measure of workload: one that tracks pilot error. Since our study was prompted by complaints received from pilots, we would also like to directly examine pilots' subjective feelings about their workload. We then decide that it would be best to collect pilots' subjective measures during the flight itself rather than tapping their memory of the flight afterwards.

This leaves us with a number of choices. The instantaneous self-assessment (ISA) technique is appealing because is promises the least amount of distraction for pilots. However, the ISA technique produces numerical ratings and presents the problem of deciding whether or not the ratings are too high or two low. The ISA technique also presents problems with pilots' interpretation of the word "workload." We consider whether or not we would have time to collect the more elaborate TLX ratings. We then consider the idea of asking pilots to fly two types of approaches: those in which reprogramming is required and those in which it is not required. This comparison would lend additional credibility to any conclusions we might draw based on workload measurement techniques that produce numerical ratings. A comparison of numerical ratings between the two types of approaches would help focus any workload problems directly on the situation in question: approaches in which pilots are asked to switch approach procedures.

The Bedford technique would be useful because it produces literal evaluations of workload levels: pilots say whether or not their workload is too high or too low. But a problem with Bedford is that it requires more time from pilots and asks them to divert their attention to a sheet of paper in order to choose a workload rating.

After some consideration, we decide that NASA TLX and Bedford subjective ratings could be comfortably elicited from pilots during the final minute of the approach, after contact with the airport has been established and workload levels relax. Based on these considerations, we decide to use three workload measurement techniques: a direct measure of pilot error; the TLX; and Bedford subjective ratings techniques. Looking at two subjective ratings of workload will allow us to compare pilots' feelings with their actual performance.

3.4.4 Preparing Materials

To assist the experimenter in collecting error data, the score sheet shown in Figure 10 was designed to allow the experimenter to quickly note any errors made by pilots during flight. A list of potential errors of interest was made and included in the score sheet. For example, pilots might stray from their assigned altitude or course, choose the wrong approach procedure in the flight management

computer, or configure the computer in an otherwise incorrect way. The experimenter can now simply observe the pilot fly and make tick marks on the score sheet along the way.

The score sheet in Figure 10 also allows the experimenter to enter Bedford and TLX ratings given by participants.

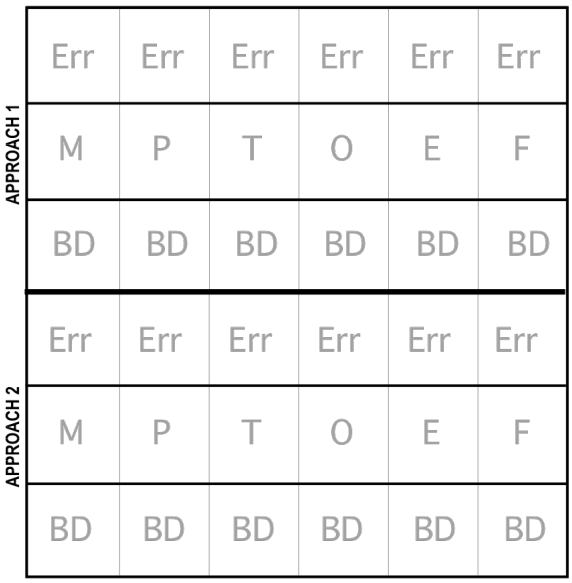


Figure 10. Experimenter score sheet for operator errors when comparing two workload approaches.

3.4.5 Collecting the Data

1. Pilots are recruited to serve as test subjects and we are left with the task of preparing to use the workload measurement techniques we have chosen.
2. The direct measure of recording errors does not require any pilot briefing since pilots are not asked to do anything other than perform the task.
3. Using the Bedford and TLX workload measures requires some additional preparation. First, pilots must be briefed on the use of both techniques. Prior to each experiment flight, each pilot is briefed on the purpose of both the Bedford and TLX techniques. The briefing should cover these items:

    a. Participant workload briefing

    - Introduce the notion of "workload" to participant
    - State that you would like to measure workload
    - Participant will be asked to rate their own workload
    - Explain that measuring workload in this way is subjective and that there are no right or wrong workload ratings. Workload is whatever the participant feels their workload to be at the time the experimenter asks.
    - Two workload measurement techniques will be used

b. Bedford

- Participant will be given the Bedford sheet
- Walk participant through process of using the Bedford sheet: making hierarchical decisions about which workload level best describes the participant's current situation
- Ask participant to ultimately provide the numerical rating that best describes their perceived workload

c. TLX

- Participant will be asked to provide six separate workload ratings
- Each of the six workload ratings asks the participant to look at workload in a different way
- Participants must give the six workload ratings verbally, using a scale of 0 to 100, by increments of 5
- Participants are asked to think of 50 as being the middle of the scale: a situation in which the participant is neither underworked nor overworked

During the flights, pilots program the assigned approach procedures into the flight management system. During half of the flights, pilots are allowed to fly the assigned approaches while the experimenter records any errors. Near the conclusion of each approach, the experimenter prompts pilots for the Bedford and TLX ratings and records them using the score sheet. During the other half of the flights, as pilots approach the destination airport ATC intervenes and asks pilots to fly a different approach. This requires pilots to focus their attention to the flight management system and reprogram the approach. The experimenter again records any errors made and near the conclusion of these approaches the experimenter collects Bedford and TLX ratings from the pilots.

3.4.6 Analyzing the Data and Drawing Conclusions

Back in the laboratory, the error data and workload ratings are compiled into a spreadsheet for analysis. Figure 11 shows the columns of data representing the errors made and the workload measures taken during the two types of approaches: those that required reprogramming and those that did not.

A statistical comparison of the data columns revealed that the approaches during which reprogramming was required results in significantly higher workload ratings for the TLX overall workload scale: $t(11)=8.789, p < .001$, as well as the Bedford scale: $t(11)=8.281, p < .001$. The number of errors committed during the two types of approaches seemed to trend toward a difference but fell short of statistical significant test.

Based on the observation that two different workload measurement techniques produced the same result, we conclude that the task of reprogramming approach procedures in flight presents a significant workload issue.

SUBJECTIVE WORKLOAD RATINGS			
Same Approach (Control)		**Different Approach**	
TLX	**Bedford**	**TLX**	**Bedford**
50	4	75	6
45	4	80	7
55	4	75	6
60	5	80	6
45	5	80	7
45	5	85	7
50	4	75	5
55	4	70	8
60	6	70	8
55	4	75	7
45	5	80	7
51.36	**4.55**	**76.82**	**6.73**

NUMBER OF ERRORS COMMITTED	
Same Approach (Control)	**Different Approach**
1	3
2	2
1	1
2	5
3	4
2	5
2	2
1	1
1	1
0	1
1	3
0	1

Figure 11. Comparative figure of subjective workload ratings and error rate.

4. References

Book, W. F. (1908). The psychology of skill. Missoula, MT: University of Montana Press.

Budescu, D.V., Zwick, R., & Rapoport, A. (1986). A Comparison of the Eigenvalue Method and the Geometric Means procedure of Ratio Scaling. Applied Psychological Measurement, 10, 69-78.

Casner, S. M. (2009). Perceived vs. Measured Effects of Advanced Cockpit Systems on Pilot Workload and Error: Are Pilots' Beliefs Misaligned With Reality? Applied Ergonomics, 40(3), 448-456.

Crawford, G., & Williams, C. (1980). Analysis of Subjective Judgment Matrices (No. Tech. Report R-2572-AF). Santa Monica, CA: The Rand Corporation.

Fairclough, S. H., Ashby, M. C., & Parkes, A. M. (1993). in-vehicle displays, visual workload and visibility evaluation. In A. G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysee & S. P. Taylor (Eds.), Vision In Vehicles - IV. Amsterdam: North-Holland.

Fisk, A. D., Derrick, W. L., & Schneider, W. (1983). The assessment of workload: Dual task methodology. Paper presented at the 27th Annual Meeting of the Human Factors Society.

Gawron, V. J. (2008). Human performance, workload, and situational awareness measures handbook (2nd ed.). Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Gray, W. D., John, B. E., & Atwood, M. E. (2002). Project Ernestine: Validating GOMS for predicting and explaining real-world task performance. Human Computer Interaction, 8(3), 237-309.

Green, R., & Flux, R. (1977). Auditory communication and workload. Paper presented at the NATO Advisory Group for Aerospace Research and Development Conference on Methods to Assess Workload.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? Human Factors, 35(4), 579-602.

Huddleston, J. H. F., & Wilson, R. V. (1971). An evaluation of the usefulness of four secondary tasks in assessing the effect of a log in simulated aircraft dynamics. Ergonomics, 14, 371-380.

Karat, C.M., Halverson, C., Horn, D. & Karat, J. (1999), Patterns of entry and correction in large vocabulary continuous speech recognition systems, CHI 99 Conference Proceedings, 568-575.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., & Wherry, R. J. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (No. 851): Army Research Institute for the Behavioral and Social Sciences.

Metalis, S. A. (1991). Heart period as a useful index of pilot workload in commercial transport aircraft. International Journal of Aviation Psychology, 1(2), 107-116.

Moray, N. (1979). Mental workload: It's theory and measurement. New York: Plenum Press.

Mulder, G. (1980). The heart of mental effort. University of Groningen.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, Kaufman, L., and Thomas, J. P. (Ed.), Handbook of perception and human performance. New York: Wiley & Sons.

Ogdon, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. Human Factors, 21(5), 529-548.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT values for predicting operator overload. Paper presented at the 32nd Annual Meeting of the Human Factors and Ergonomics Society.

Roscoe, A. H. (1984). Assessing pilot workload in flight: Flight test techniques (No. AGARD-CP-373). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development (AGARD).

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV, and respiration? Biological Psychology, 34, 259-288.

Rueb, J. D., Vidulich, M. A., & Hassoun, J. A. (1994). Use of workload redlines: A KC-135 crew-reduction application. International Journal of Aviation Psychology, 4, 47-64.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. Human Factors, 48(2), 381-391.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. Human Factors, 29(2), 171-182.

Vidulich, M. A., & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. Applied Ergonomics, 17, 291-296.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. Human Factors, 30, 111-120.

5. Bibliography

The following articles are the result of a literature search on the topic of human workload measurement and management. We limited our bibliography to contain only those articles that have been published in peer-reviewed scientific journals or books. In a limited number of cases we have included articles published in conferences and meetings or as technical reports when these articles have received consistent citations in other peer-reviewed work.

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. Human Factors, 29(2), 161-170.

Andre, A. D., Heers, S. T., & Cashion, P. A. (1995). Effects of workload preview on task scheduling during simulated instrument flight. International Journal of Aviation Psychology, 5(1), 5-23.

Backs, R. W., Ryan, A. M., & Wilson, G. F. (1994). Psychophysiological measures of workload during continuous manual performance. Human Factors, 36, 514-531.

Backs, R. W. (1995). Going beyond heart rate: Autonomic space and cardiovascular assessment of mental workload. International Journal of Aviation Psychology, 5(1), 25-48.

Backs, R. W., Lenneman, J. K., & Sicard, J. L. (1999). The use of autonomic components to improve cardiovascular assessment of mental workload in flight simulation. International Journal of Aviation Psychology, 9(1), 33-47.

Becker, A. B., Warm, J. S., & Dember, W. N. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. International Journal of Aviation Psychology, 5(1), 49-62.

Book, W. F. (1908). The psychology of skill. Missoula, MT: University of Montana Press.

Borg, C. G. (1978). Subjective aspects of physical and mental load. Ergonomics, 21, 215-220.

Bortolussi, M. R., Kantowitz, B. H., & Hart, S. G. (1986). Measuring pilot workload in a motion base trainer. Applied Ergonomics, 17(4), 278-283.

Braby, C. D., Harris, D., & Muir, H. C. (1993). A psycholophysiological approach to the assessment of work underload. Ergonomics, 36, 1035-1042.

Brookings, J., Wilson, G. F., & Swain, C. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology, 42, 361-378.

Brown, I. D., & Poulton, E. C. (1961). Measuring the "spare mental capacity" of car drivers by a subsidiary auditory task. Ergonomics, 4, 35-40.

Brown, I. D. (1978). Dual task methods of assessing work-load. Ergonomics, 21(3), 221-224.

Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: effects of mental workload and event structure. J Exp Psychol Hum Percept Perform, 28(3), 600-615.

Budescu, D.V., Zwick, R., & Rapoport, A. (1986). A Comparison of the Eigenvalue Method and the Geometric Means procedure of Ratio Scaling. Applied Psychological Measurement, 10, 69-78.

Cain, B. (2007). A review of the mental workload literature (No. RTO-TR-HFM-121). Toronto: Defence Research and Development Canada.

Casali, J. G., & Wierwille, W. W. (1983). A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 25(6), 623-641.

Casner, S. M. (2009). Perceived vs. Measured Effects of Advanced Cockpit Systems on Pilot Workload and Error: Are Pilots' Beliefs Misaligned With Reality? Applied Ergonomics, 40(3), 448-456.

Charlton, S. G. (1996). Mental workload test and evaluation. In T.G. O'Brien & S.G. Charlton (Eds.), Handbook of Human Factors Testing and Evaluation. Mahweh, NJ: Lawrence Erlbaum.

Charlton, S. G. (2009). Driving while conversing: cell phones that distract and passengers who react. Accid Anal Prev, 41(1), 160-173.

Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or operational workload with performance-measurement methods. Human Factors, 21(5), 515-528.

Comens, P., Reed, D., & Mette, M. (1987). Physiologic responses of pilots flying high-performance aircraft. Aviation Space and Environmental Medicine, 58, 205-210.

Crawford, G., & Williams, C. (1980). Analysis of Subjective Judgment Matrices (No. Tech. Report R-2572-AF). Santa Monica, CA: The Rand Corporation.

Damos, D. (1985). The relation between the Type A behavior pattern, pacing, and subjective workload under single and dual-task conditions. Human Factors, 27(6), 675-680.

Egelund, N. (1982). Spectral analysis of heart rate variability as an indicator of driver fatigue. Ergonomics, 25, 663-672.

Elmenhorst, E.-M., Vejvoda, M., Maass, H., Wenzel, J., Plath, G., Schubert, E., & Basner, M. (2009). Pilot workload during approaches: Comparison of simulated standard and noise-abatement profiles. Aviation Space and Environmental Medicine, 80(4), 364-370.

Endsley, M. R., & Kiris, E. O. (1993). Situation awareness and workload: Flip sides of the same coin. Paper presented at the Seventh International Symposium on Aviation Psychology, Columbus, OH.

Endsley, M. R., & Kaber, D. B. (1999). Level of autoomation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42(3), 462-492.

Fairclough, S. H., Ashby, M. C., & Parkes, A. M. (1993). in-vehicle displays, visual workload and visibility evaluation. In A. G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruysee & S. P. Taylor (Eds.), Vision In Vehicles - IV. Amsterdam: North-Holland.

Farmer, E., & Brownson, A. (2003). Review of workload measurement, analysis and interpretation methods (No. CARE-Integra-TRS-130-02-WP2): European Organisation for the Safety of Air Navigation (EUROCONTROL).

Fisk, A. D., Derrick, W. L., & Schneider, W. (1983). The assessment of workload: Dual task methodology. Paper presented at the 27th Annual Meeting of the Human Factors Society.

Furedy, J. J. (1987). Beyond heart rate in the cardiac physiological assessment of mental effort: The T-wave amplitude coomponent of the electrocardiogram. Human Factors, 29(2), 183-194.

Gabriel, R. F., & Burrows, A. A. (1968). Improving time-sharing performance of pilots through training. Human Factors, 10, 33-40.

Gawron, V. J. (2008). Human performance, workload, and situational awareness measures handbook (2nd ed.). Boca Raton, Florida: CRC Press, Taylor & Francis Group.

Gopher, D., & Navon, D. (1980). How is performance limited: Testing the notion of central capacity. Acta Psychologica, 46, 161-180.

Gopher, D., & Donchin, E. (1986). Workload: an examination of the concept. In L. Kaufman & J. Thomas K. Boff (Ed.), Handbook of perception and human performance. New York: Wiley & Sons, Inc.

Gopher, D. & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? Human Factors, 26(5), 519-532.

Gould, K. S., Roed, B. K., Saus, E.-R., Koefoed, V. F., Bridger, R. S., & Moen, B. E. (2009). Effects of navigation method on workload and performance in simulated high-speed ship navigation. Applied Ergonomics, 40, 103-114.

Gray, W. D., John, B. E., & Atwood, M. E. (2002). Project Ernestine: Validating GOMS for predicting and explaining real-world task performance. Human Computer Interaction, 8(3), 237-309.

Green, R., & Flux, R. (1977). Auditory communication and workload. Paper presented at the NATO Advisory Group for Aerospace Research and Development Conference on Methods to Assess Workload.

Hamilton, D. B., Bierbaum, C. R., & Filford, L. A. (1991). Task analysis/workload (TAWL) User's guide - version 4.0 (No. ASI 690-330-90). Fort Rucker, Alabama: Anacapa Sciences.

Hancock, P. A., Meshkati, N., & Robertson, M. M. (1985). Physiological reflections of mental workload. Aviation Space and Environmental Medicine, 56(11), 1110-1114.

Hancock, P.A., & Chignell, M.H. (1988). Mental workload dynamics in adaptive interface design. IEEE Transactions on Systems, Man, and Cybernetics, 18(4), 647-659.

Hancock, P. A., & Caird, J. K. (1993). Experimental evaluation of a model of mental workload. Human Factors, 35(3), 413-430.

Hancock, P. A., Manning, C. M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. International Journal of Aviation Psychology, 5(1), 63-86.

Harris, W. C., Hancock, P. A., Arthur, E. J., & Caird, J. K. (1995). Performance, workload, and fatigue changes associated with automation. International Journal of Aviation Psychology, 5(2), 169-185.

Hart, S.G. (1975). Time estimation as a secondary task to measure workload. Paper presented at the 11th Annual Conference on Manual Control.

Hart, S.G. (1978). Subjective time estimation as an index of workload. Paper presented at the Symposium on Man-machine interface: advances in workload study.

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. Human Factors, 26(5), 545-556.

Hart, S.G. (1987). Research papers and publications (1981-1987): Workload research program (Technical Memorandum No. N88-12924). Moffett Field, CA: Ames Research Center.

Hart, S. G., & Staveland, L. (1988). Development of the NASA task load index (TLX): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 239-250). Amsterdam: North Holland.

Haskell, B. E., & Reid, G. B. (1987). The subjective perception of workload in low-time private pilots: A preliminary study. Aviation Space and Environmental Medicine, 58, 1230-1232.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? Human Factors, 35(4), 579-602.

Hendy, K.C. (1995). Situation Awareness and Workload: Birds of a Feather? Paper presented at the AGARD AMP Symposium on 'Situational Awareness: Limitations and Enhancements in the Aviation Environment.

Hendy, K.C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. Human Factors, 39(1), 30-47.

Hering, H., & Coatleven, G. (1996). ERGO (version 2) for Instantaneous Self Assessment of workload in a real time ATC simulation environment (No. EEC Report No10/96). Bruxells, Belgium: Eurocontrol Agency.

Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base simulator. Human Factors, 21, 129-143.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34(4), 429-439.

Hoh, R. H., Smith, J. C., & Hinton, D. A. (1987). The effects of display and autopilot functions on pilot workload for single pilot instrument flight (SPIFR) operations. (No. NASA Contractor Report 4073). Langley, VA: NASA Langley Research Center.

Huddleston, J. H. F., & Wilson, R. V. (1971). An evaluation of the usefulness of four secondary tasks in assessing the effect of a log in simulated aircraft dynamics. Ergonomics, 14, 371-380.

Isreal, J. B., Wickens, C. D., Chesney, G., & Donchin, E. (1980). The event related brain potential as a selective index of display monitoring load. Human Factors, 22, 211-224.

Jordan, N. (1963). Allocation of functions between man and machines in automated systems. Journal of Applied Psychology, 47(3), 161-165.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and physiological state: A review of its validity as a workload index. Biological Psychology, 34, 237-258.

Kalsbeek, J. W. H., & Sykes, R. N. (1967). Objective measurement of mental load. In A. F. Sanders (Ed.), Attention and performance. Amsterdam: North-Holland.

Kantowitz, B. H. (1992). Selecting measures for human factors research. Human Factors, 34, 387-398.

Kantowitz, B. H. & Casper, P. A. (1988). Human workload in aviation. In E. L. Wiener and D. C. Nagel (Ed.), Human Factors In Aviation (pp. 157-187). San Diego: Academic Press.

Karat, C.M., Halverson, C., Horn, D. & Karat, J. (1999), Patterns of entry and correction in large vocabulary continuous speech recognition systems, CHI 99 Conference Proceedings, 568-575.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. Human Factors, 35(2), 221-242.

Kirwin, B., and Ainsworth, L. K. (1992). A guide to task analysis. London: Taylor & Francis.

Kitay, D. (1978). Crew workload in the air carrier cockpit: Air Line Pilots Association.

Knowles, W. B. (1963). Operating loading tasks. Human Factors, 5, 151-161.

Kramer, A. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), Multiple task performance (pp. 279-328). London: Taylor & Francis.

Kramer, A. F., Sirevaag, E. J., & Braune, R. A. (1987). A psychophysical assessment of operator workload during simulated flight missions. Human Factors, 29, 145-160.

Kramer, A. F., Trejo, D., & Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. Biological Psychology, 39, 83-101.

Krol, J. P. (1971). Variations in ATC workload as a function of variations in cockpit workload. Ergonomics, 14, 585-590.

Lamoureux, T. (1999). The influence of aircraft proximity on the subjective mental workload of controllers in the air traffic control task. Ergonomics, 42(11), 1482-1491.

Laudeman, I. V., & Palmer, E. A. (1995). Quantitative measurement of observed workload in the analysis of aircrew performance. International Journal of Aviation Psychology, 5(2), 187-197.

Leplat, J., & Welford, A. T. (1978). Special issue: Symposium on mental workload. Ergonomics, 21, 141-233.

Liu, Y., & Wickens, C.D. (1993). Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. Ergonomics, 37(11), 1843-1854.

Luximon, A. & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. Ergonomics, 44(3), 229-243.

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., & Wherry, R. J. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (No. 851): Army Research Institute for the Behavioral and Social Sciences.

Matthews, M. L. (1986). The influence of visual workload history on visual performance. Human Factors, 28(6), 623-632.

Metalis, S. A. (1991). Heart period as a useful index of pilot workload in commercial transport aircraft. International Journal of Aviation Psychology, 1(2), 107-116.

Michon, J. A. (1966). Tapping regularity as a measure of perceptual motor load. Ergonomics, 9(5), 401-412.

Moray, N. (1979). Mental workload: It's theory and measurement. New York: Plenum Press.

Moray, N., Johanssen, J., Pew, R. D., Rasmussen, J., Sanders, A. F., and Wickens, C. D. (1979). Report of the experimental psychology group. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York: Plenum.

Moray, N. (1982). Subjective mental workload. Human Factors, 24(1), 25-40.

Moray, N., Dessouky, M.I., Kijowski, B.A., & Adapathaya, R. (1991). Strategic behavior, workload and performance in task scheduling. Human Factors, 33(6), 607-629.

Moroney, W. F., Biers, D. W., & Eggemeier, F. T. (1995). Some measurement and methodological considerations in the application of subjective workload measurement techniques. International Journal of Aviation Psychology, 5(1), 87-106.

Morris, C. H., & Leung, Y. K. (2006). Pilot mental workload: How well do pilots really perform? Ergonomics, 49(15), 1581-1596.

Muckler, F. A., & Seven, S. A. (1992). Selecting performance measures: objective versus subjective measurement. Human Factors, 34, 441-456.

Mulder, G. (1980). The heart of mental effort. University of Groningen.

Mulder, L. J. M. (1992). Measurement and analysis of heart rate and respiration for use in applied environments. Biological Psychology, 34, 205-236.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in assessment of perceived mental workload. Human Factors, 33(1), 17-33.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, Kaufman, L., and Thomas, J. P. (Ed.), Handbook of perception and human performance. New York: Wiley & Sons.

Ogdon, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. Human Factors, 21(5), 529-548.

Oman, C. M., Kendra, A. J., Hayashi, M., Stearns, M. J., & Burki-Cohen, J. (2001). Vertical navigation displays: Pilot performance and workload during simulated constant angle of descent GPS approaches. International Journal of Aviation Psychology, 11(1), 15-31.

Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. In D. J. Garland and J. A. Wise (Ed.), Human factors and advanced aviation technologies (pp. 147-158). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced complacency. International Journal of Aviation Psychology, 3(1), 1-23.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39, 230-253.

Parasuraman, R., & Rovira, E. (2005). Workload Modeling and Workload Management: Recent Theoretical Developments (No. ARL-CR-0562). Aberdeen Proving Ground: Army Research Laboratory.

Perala, C.H., & Sterling, B.S. (2007). Galvanic Skin Response as a Measure of Soldier Stress (Final No. ARL-TR-4114). Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Perry, C. M., Sheik-Nainar, M. A., Segall, N., Ma, R., & Kaber, D. B. (2006). Effects of physical workload on cognitive task performance and situation awareness. Theoretical Issues in Ergonomics Science, 9(2), 95-113.

Porges, S. W., & Byrne, E. A. (1992). Research emthods for measurement of heart rate and respiration. Biological Psychology, 34, 93-130.

Porterfield, D. H. (1997). Evaluating controller communication time as a measure of workload. International Journal of Aviation Psychology, 7(2), 171-182.

Raby, M., & Wickens, C.D. (1994). Strategic workload management and decision biases in aviation. International Journal of Aviation Psychology, 4(3), 211-240.

Rehmann, J. T., Stein, E. S., & Rosenberg, B. L. (1983). Subjective pilot workload assessment. Human Factors, 25(3), 297-307.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT values for predicting operator overload. Paper presented at the 32nd Annual Meeting of the Human Factors and Ergonomics Society.

Reid, G.B., Potter, S.S., & Bressler, J.R (1989). Subjective workload assessment technique (SWAT): a user's guide (No. AAMRL-TR-89-023). WPAFB, Ohio: Armstrong Aerospace Medical Research Laboratory.

Roscoe, A. H. (1984). Assessing pilot workload in flight: Flight test techniques (No. AGARD-CP-373). Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research and Development (AGARD).

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV, and respiration? Biological Psychology, 34, 259-288.

Roscoe, A.H., & Ellis, G.A. (1990). A subjective rating scale for assessing pilot workload in flight: a decase of practical use (No. Technical Report TR 90019). Farnborough, UK: Royal Aerospace Establishment.

Rouse, W.B, Edwards, S.L., & Hammer, J.M. (1993). Modeling the dynamics of mental workload and human performance in complex systems. IEEE Transactions on Systems, Man, and Cybernetics, 23(6), 1-10.

Rueb, J. D., Vidulich, M. A., & Hassoun, J. A. (1994). Use of workload redlines: A KC-135 crew-reduction application. International Journal of Aviation Psychology, 4, 47-64.

Sarno, K. J., & Wickens, C. D. (1995). Role of multiple resources in predicting time-sharing efficiency: Evaluation of three workload models in a multiple-task setting. International Journal of Aviation Psychology, 5(1), 107-130.

Savage, R. E., Wierwille, W. W., & Cordes, R. E. (1978). Evaluating the sensitivity of various measures of operator workload using random digits as a secondary task. Human Factors, 20(6), 649-654.

Schouten, J. F., Kalsbeek, J. W. H., & Leopold, F. F. (1962). On the evaluation of perceptual and mental load. Ergonomics, 5, 251-260.

Sheridan, T. B. (1980). Computer control and human alienation. Technology Review, 10, 61-73.

Shinohara, K., Miura, T., & Usui, S. (2002). Tapping task as an index of mental workload in a time-sharing task. Japanese Psychological Research, 44(3).

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced complacency: Development of the complacency-potential rating scale. International Journal of Aviation Psychology, 3(2), 111-122.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. Ergonomics, 36, 1121-1140.

Sohn, S. Y., & Jo, Y. K. (2003). A study on the student pilot's mental workload due to personality types of both instructor and student. Ergonomics, 46(15), 1566-1577.

Stein, E. S., & Rosenberg, B. L. (1982). The measurement of pilot workload (No. DOT/FAA/CT-82/83). Washington, D.C.: Federal Aviation Administration.

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. Human Factors, 48(2), 381-391.

Tattersall, A. J. & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics, 39(5), 740-748.

Tsang, P. S., & Vidulich, M. A. (1994). The roles of immediacy and redundancy in relative subjective workload assessment. Human Factors, 36(3), 503-513.

Tsang, P. S., & Velasquez, V. L. (1996). Diagnosticity and multidimensional subjective workload rating. Ergonomics, 39(3), 358-381.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. Biological Psychology, 42, 323-342.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral analysis of sinus arrhythmia: A measure of mental effort. Human Factors, 29(2), 171-182.

Vidulich, M. A., & Tsang, P. S. (1986). Techniques of subjective workload assessment: A comparison of SWAT and NASA-bipolar methods. Ergonomics, 29, 1385-1398.

Vidulich, M. A., & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance: Caveats for the use of subjective assessments. Applied Ergonomics, 17, 291-296.

Vidulich, M. A. (1989). The use of judgment matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. Paper presented at the Human Factors and Ergonomics Society Annual Meeting.

Vidulich, M. A., Ward, G. F., & Schueren, J. (1991). Using subjective workload dominance (SWORD) technique for predictive workload assessment. Human Factors, 33, 677-692.

Vidulich, M. A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and physiological measures of situation awareness. Aviation Space and Environmental Medicine, 65(5), A7-A12.

Vidulich, M.A., & Tsang, P.S. (1987, 1987). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. Paper presented at the Human Factors Society 31st Annual Meeting.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman, and Mouloua, M. (Ed.), Automation and Human Performance (pp. 183-200). Mahwah, NJ: Lawrence Erlbaum Associates.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1998). Workload and vigilance. Paper presented at the Human Factors and Ergonomics Society 42nd Annual Meeting.

Welford, A. T. (1978). Mental workload as a function of demand, capacity, strategy, and skill. Ergonomics, 21, 151-167.

Wickens, C. D., & Kessel, C. (1979). The effect of participatory mode and task workload on the detection of dynamic system failures. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 24-34.

Wickens, C. D., Hyman, F., Dellinger, J., Taylor, H., & Meador, M. (1986). The Sternberg memory search task as an index of pilot workload. Ergonomics, 29, 1371-1383.

Wickes, C. D. (2002). Situation awareness and workload in aviation. Current Directions in Psychological Science, 11(4), 128-133.

Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: Promises and problems. Ergonomics, 23, 995-1011.

Wierwille, W. W. (1979). Physiological measure of aircrew mental workload. Human Factors, 21(5), 575-593.

Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 25(1), 1-16.

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. Human Factors, 27(5), 489-502.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35(2), 263-281.

Willigies, R. C., & Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.

Wilson, G. (2001). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. International Journal of Aviation Psychology, 12(1), 3-18.

Wilson, G. F. (1993). Air-to-ground training missions: a psychophysiological workload analysis. Ergonomics, 36, 1071-1087.

Xie, B. (2000). Prediction of mental workload in single and multiple tasks environments. International Journal of Cognitive Ergonomics, 4(3), 213-242.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. Human Factors, 30, 111-120.